

<p>Nama: <b>Ibnu Fajar Setiawan</b></p> <p>NIM: <b>065002000006</b></p>	 <p><b>Praktikum Data Warehouse</b></p>	<h1>MODUL 11</h1> <p>Nama Dosen: <b>Ir. Teddy Siswanto, MMSi</b></p> <p>Nama Asisten Laboratorium:</p> <ul style="list-style-type: none"><li>1. Azhar Rizki Zulma <b>065001900001</b></li><li>2. Nadiya Amanda Rizkania <b>064001900003</b></li></ul>
---------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## Proyek Akhir Praktikum III

### 1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

### Cleansing Data

Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah-langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:



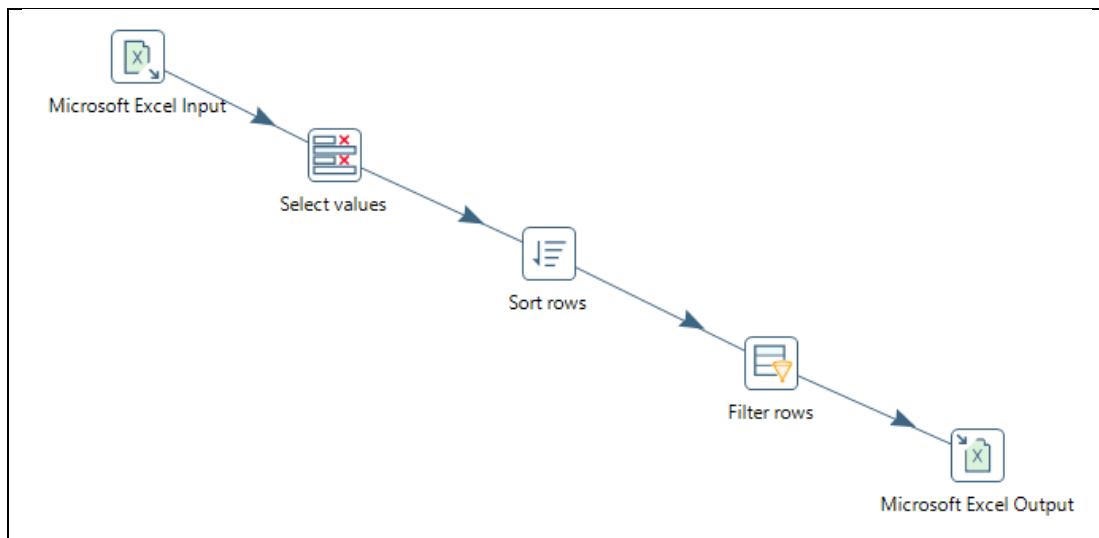
## 1. Data Preparation

Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.

	Name	Date modified	Type	Size
ss	Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
ds	Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
nts	Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
ts	Cabai Merah Keriting September 2018	14/11/2020 20:45	Microsoft Excel C...	10 KB
tp	data cabai keriting merah 2018 (jan ags)	01/05/2021 12:10	Microsoft Excel 97...	108 KB

## 2. Excel Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2020.ktr***



- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

The screenshot shows two windows of the Microsoft Excel Input configuration interface.

**Top Window (Files Tab):**

- Step name:** Microsoft Excel Input
- Spread sheet type (engine):** Excel 97-2003 XLS (JXL)
- Selected files:** C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Data Warehouse\Praktikum Video Cleansing and Filter
- Accept filenames from previous steps:**
- Accept filenames from previous step:**
- Step to read filenames from:** (dropdown menu)
- Field in the input to use as filename:** (dropdown menu)

**Bottom Window (Sheets Tab):**

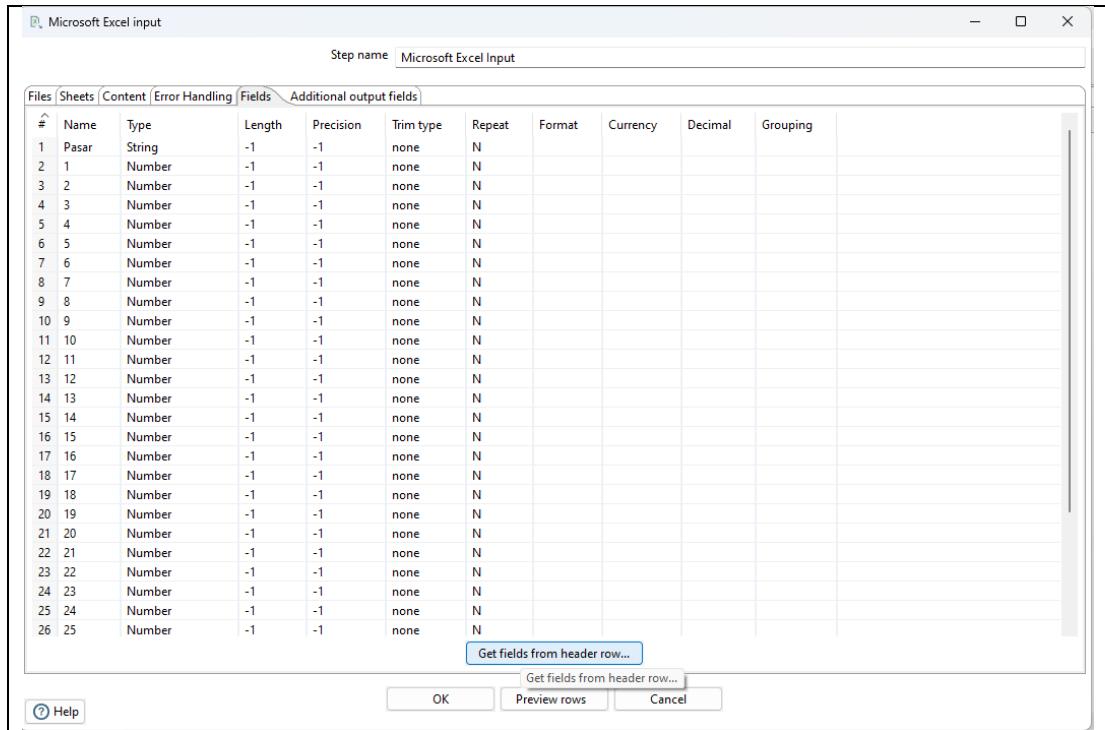
- Step name:** Microsoft Excel Input
- List of sheets to read:**

#	Sheet name	Start row	Start column
1	agustus 2018	0	0

- Enter list dialog:**

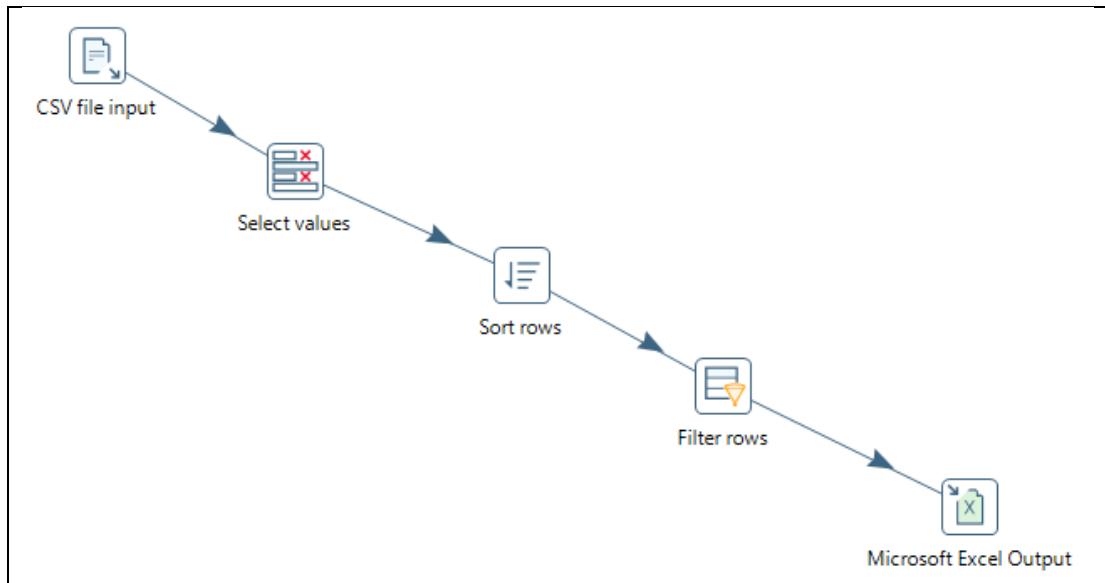
  - Available items:** januari 2018, Februari 2018, maret 2018, april 2018, mei 2018, juni 2018, juli 2018
  - Your selection:** agustus 2018
  - Buttons:** >, <, <<, OK, Cancel



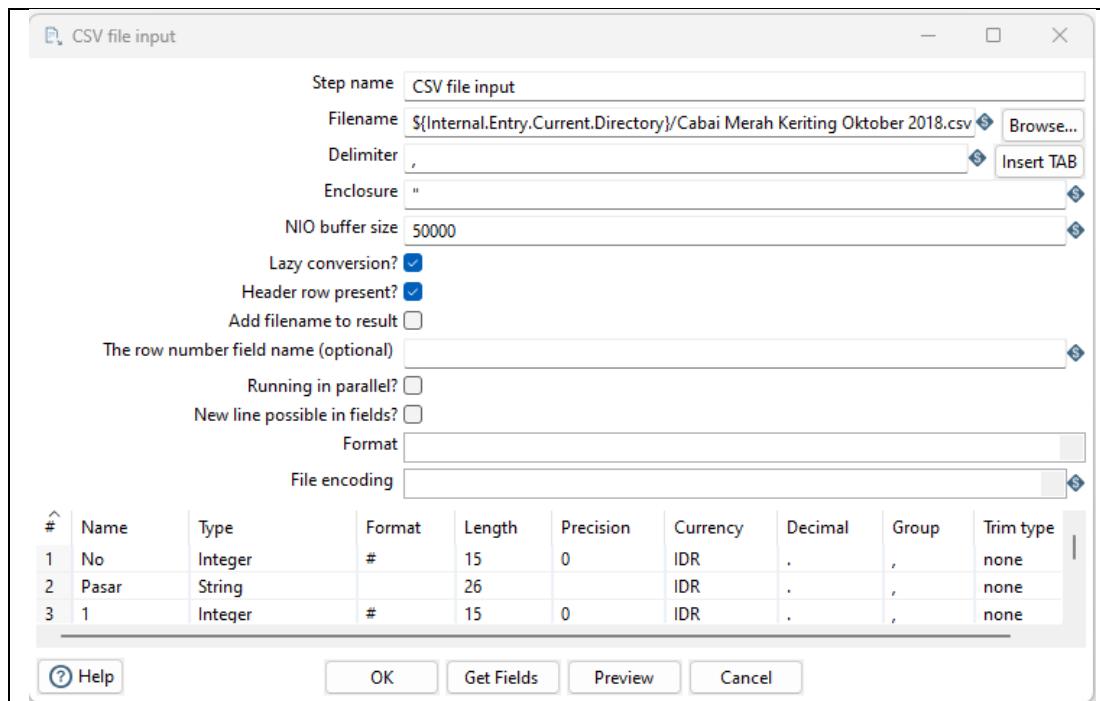


### 3. CSV Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format **Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr** seperti **Cleansing Cabai Merah Keriting Oktober 2020.ktr**

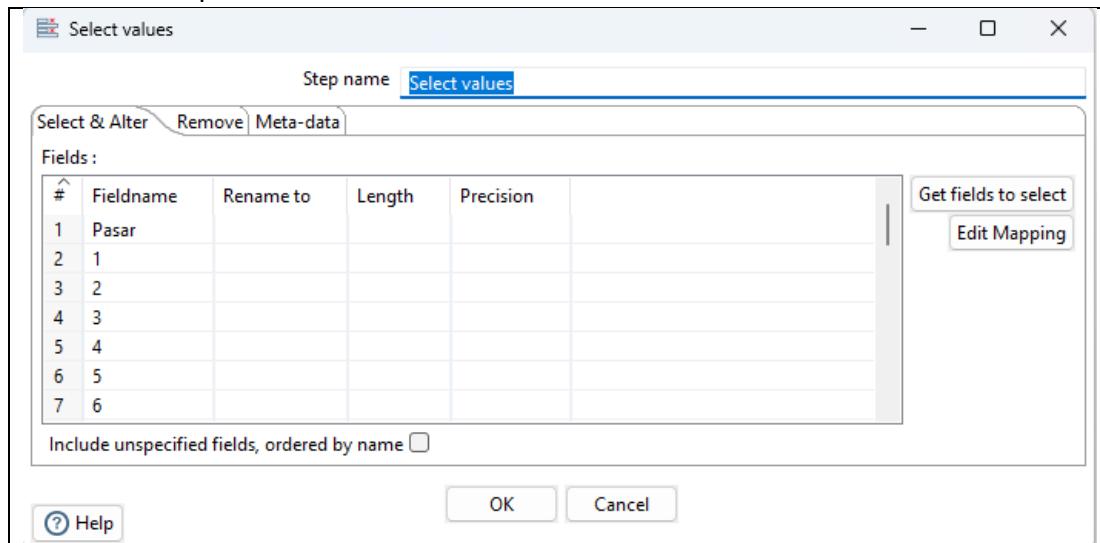


- Berikut ini adalah konfigurasi pada step CSV File Input.

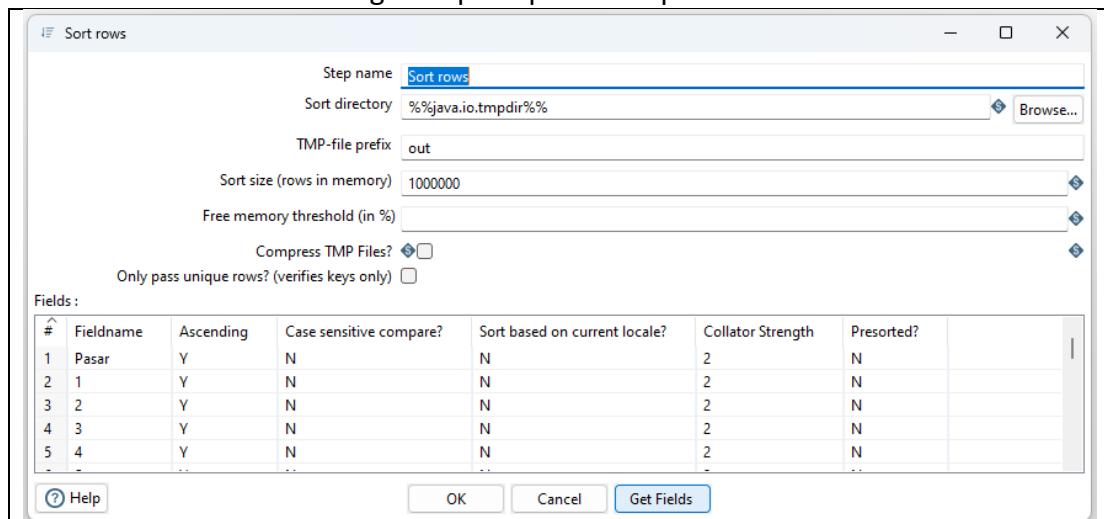


#### 4. Data Filtering

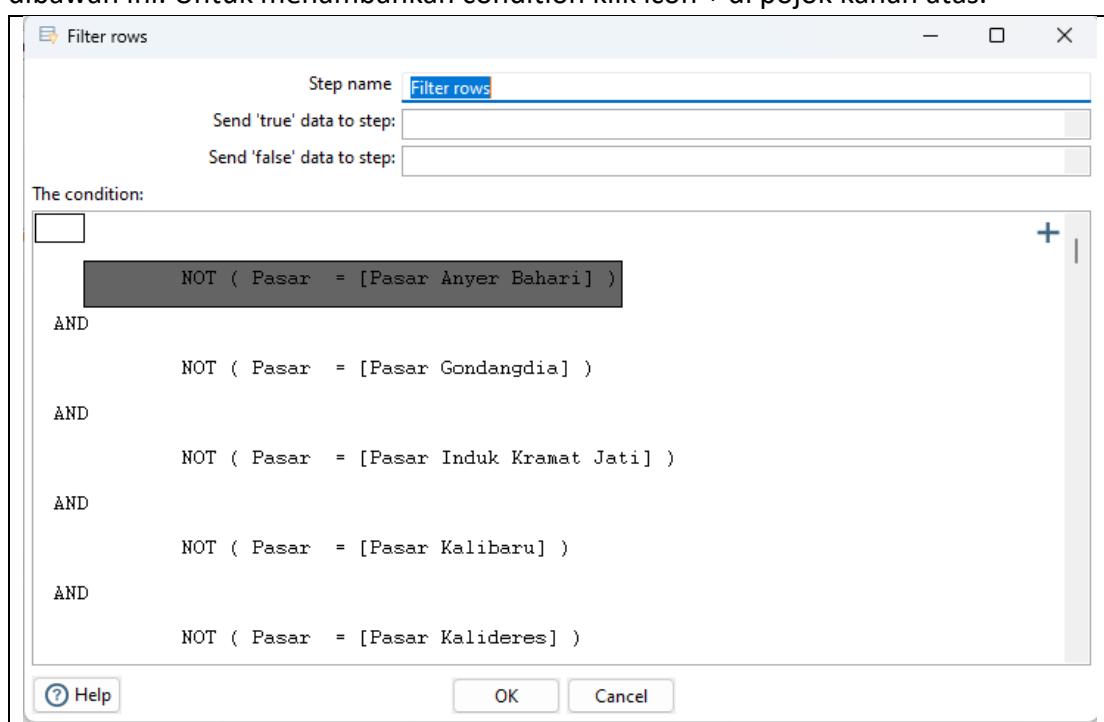
- Step selanjutnya adalah melakukan select value yang berguna untuk memilih kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.

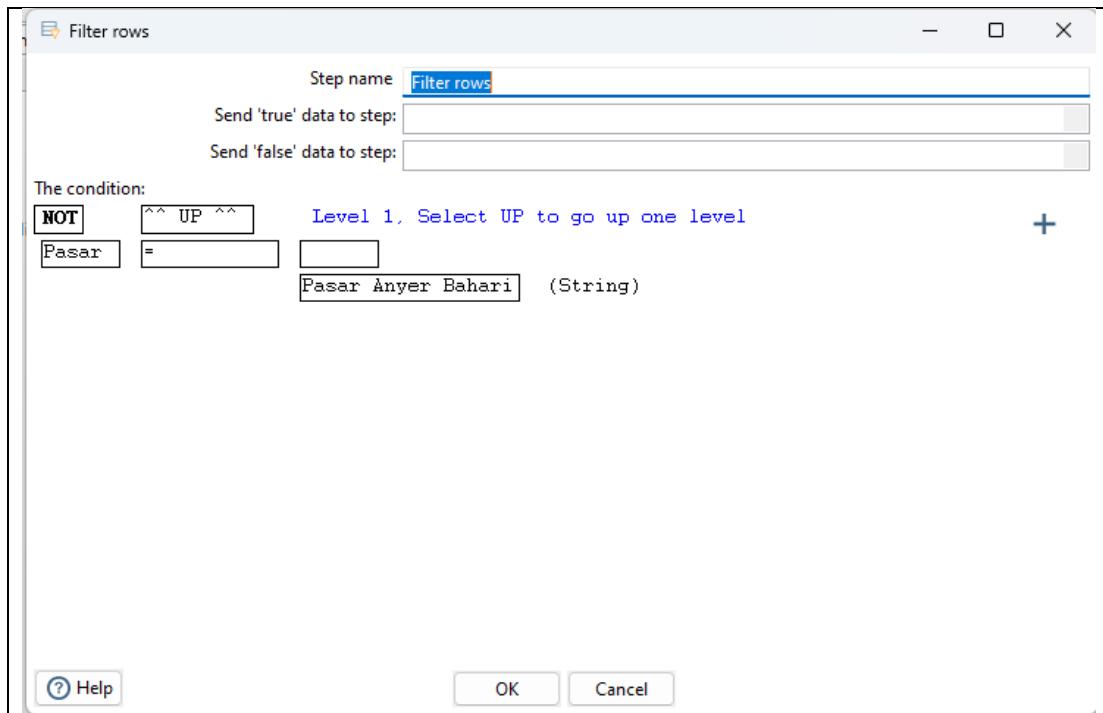


- Kemudian lakukan ascending data pada proses step Sort Rows



- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.





- Berikut List Data Nama Pasar yang mau dibuang (2020)

```
NOT ( Pasar = [Pasar Anyer Bahari] )
AND
NOT ( Pasar = [Pasar Gondangdia] )
AND
NOT ( Pasar = [Pasar Induk Kramat Jati] )
AND
NOT ( Pasar = [Pasar Kalibaru] )
AND
NOT ( Pasar = [Pasar Kalideres] )
AND
NOT ( Pasar = [Pasar Klender SS] )
AND
NOT ( Pasar = [Pasar Koja Baru] )
AND
NOT ( Pasar = [Pasar Rawa Sanggun] )
AND
NOT ( Pasar = [Pasar Pramuka] )
AND
NOT ( Pasar = [Pasar Pluit] )
AND
NOT ( Pasar = [Pasar Pulo Gadung] )
AND
NOT ( Pasar = [Pasar Pesanggrahan] )
AND
NOT ( Pasar = [Pasar Glodok] )
AND
NOT ( Pasar = [Pasar Matraman KK] )
AND
NOT ( Pasar = [Pasar Jatinegara] )
```



## 5. Excel Output

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama **Cabai Merah Keriting [Bulan] [Tahun].xls** seperti **Cabai Merah Keriting Agustus 2020.xls** (File & Sheet)

The image shows two overlapping windows for configuring Microsoft Excel output. The top window is titled 'Microsoft Excel output' and has a 'Step name' of 'Microsoft Excel Output'. It includes fields for 'Filename' (C:\Users\Azhar Rizky Zulma\OneDrive\\$), 'Extension' (xls), and various options like 'Include stepnr in filename?' and 'Add filenames to result'. The bottom window is also titled 'Microsoft Excel output' with the same 'Step name'. It displays a table of fields with columns '#', 'Name', 'Type', and 'Format'. The data is as follows:

#	Name	Type	Format
1	Pasar	String	
2	1	Integer	
3	2	Integer	
4	3	Integer	
5	4	Integer	
6	5	Integer	



## 2. Alat dan Bahan

Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

## 3. Elemen Kompetensi

a. Tugas Akhir Praktikum III – Screenshot Struktur yang sudah berhasil berjalan

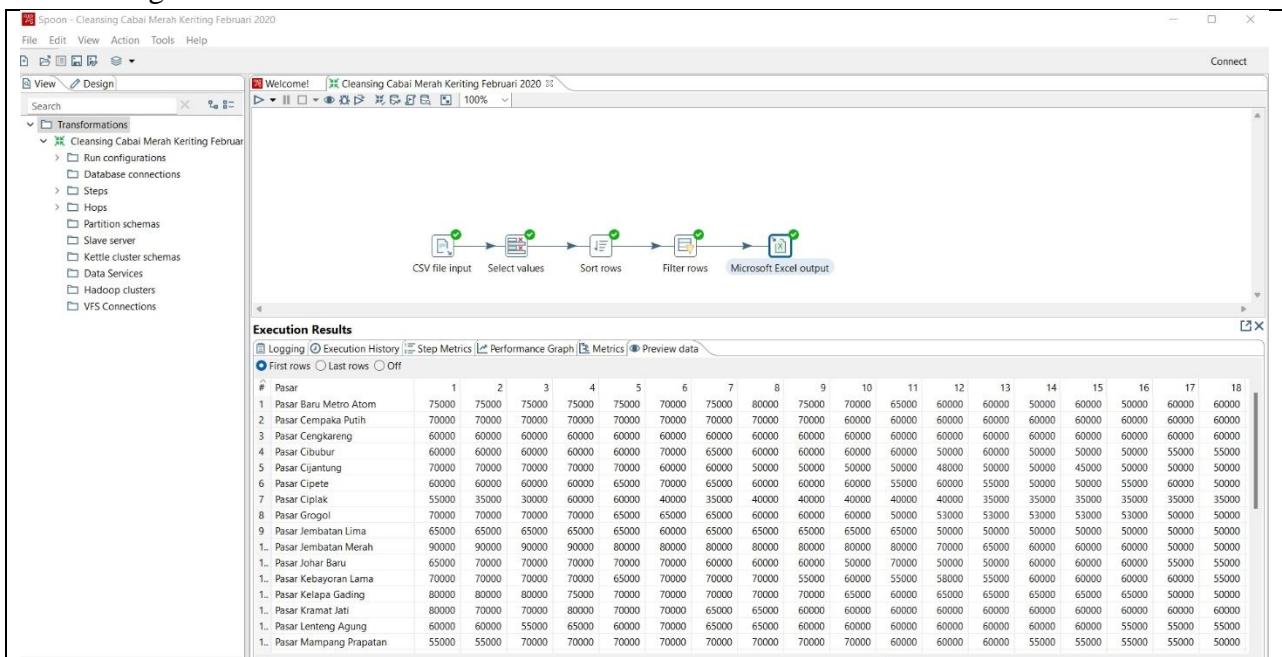
### 1. Cleansing Januari 2020

The screenshot shows the Spoon Pentaho Data Integration interface. On the left, the 'Design' view displays a workflow: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. The 'Execution Results' pane shows the first 16 rows of the cleaned data, which lists various markets (Pasar) and their corresponding values. Below this is a screenshot of an Excel spreadsheet titled 'Cabai Merah Keriting Januari 2020.xlsx' containing the same data.

Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Pasar																
1 Pasar Baru Metro Atom	45000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	70000	65000
2 Pasar Cempaka Putih	40000	40000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	70000	70000
3 Pasar Cengkareng	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	60000	60000	60000
4 Pasar Citubur	40000	60000	50000	50000	50000	50000	50000	70000	50000	50000	60000	60000	60000	55000	60000	55000
5 Pasar Cijantung	50000	40000	65000	65000	65000	60000	60000	70000	70000	70000	70000	70000	70000	70000	70000	70000
6 Pasar Cipelet	45000	45000	50000	50000	55000	60000	65000	65000	65000	65000	65000	65000	65000	60000	65000	65000
7 Pasar Ciplak	30000	30000	30000	30000	30000	30000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
8 Pasar Grogol	40000	40000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	65000	65000	65000



## 2. Cleansing Februari 2020



The screenshot shows the Kettle Data Integration tool interface. On the left, the 'Transformations' tree view is expanded to show the 'Cleansing Cabai Merah Keriting Februar' project. The main workspace displays a data flow diagram with five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the diagram, the 'Execution Results' tab is active, showing a preview of the data. The data table has 18 columns and 30 rows, listing various market names (Pasar) and their corresponding values. The Excel application window below shows the full dataset with many empty cells, indicating missing or null data.

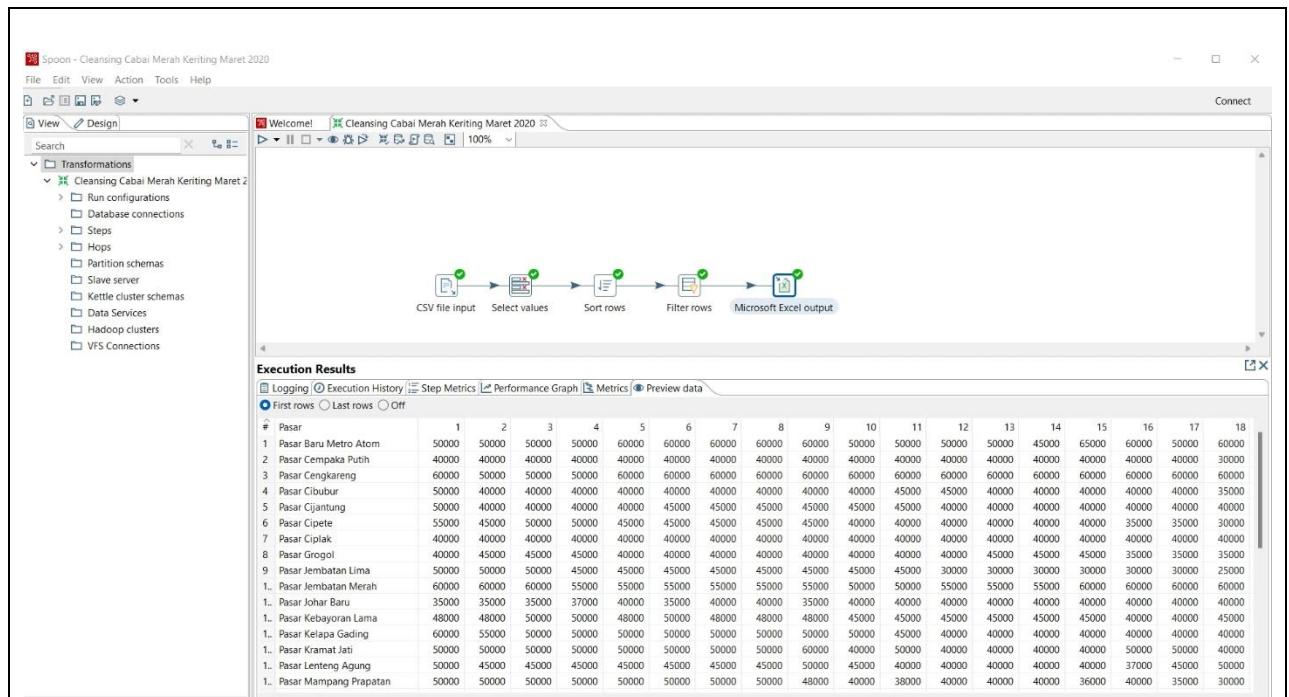
Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Pasar																		
1. Pasar Baru Metro Atom	75000	75000	75000	75000	75000	70000	75000	80000	75000	70000	65000	60000	60000	50000	60000	50000	60000	60000
2. Pasar Cempaka Putih	70000	70000	70000	70000	70000	70000	70000	70000	70000	60000	60000	60000	60000	60000	60000	60000	60000	60000
3. Pasar Cengkareng	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
4. Pasar Cibubur	60000	60000	60000	60000	60000	60000	70000	65000	60000	60000	60000	60000	55000	60000	55000	50000	50000	55000
5. Pasar Cijantung	70000	70000	70000	70000	70000	60000	60000	50000	50000	50000	48000	50000	50000	45000	50000	50000	50000	50000
6. Pasar Cipete	60000	60000	60000	60000	60000	60000	70000	65000	60000	60000	55000	60000	50000	50000	50000	50000	50000	50000
7. Pasar Ciplak	55000	35000	30000	60000	60000	40000	35000	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000
8. Pasar Grogol	70000	70000	70000	70000	65000	65000	65000	60000	65000	65000	65000	50000	53000	53000	53000	53000	53000	50000
9. pasar Jembatan Lima	65000	65000	65000	65000	65000	60000	65000	65000	65000	65000	65000	50000	50000	50000	50000	50000	50000	50000
1. Pasar Jembatan Merah	90000	90000	90000	80000	80000	80000	80000	80000	80000	80000	70000	65000	60000	60000	60000	60000	60000	60000
1. Pasar Johar Baru	65000	70000	70000	70000	70000	60000	60000	60000	60000	60000	50000	50000	50000	60000	60000	55000	55000	55000
1. Pasar Kebayoran Lama	70000	70000	70000	70000	65000	70000	70000	70000	55000	60000	55000	55000	60000	60000	60000	60000	60000	55000
1. Pasar Kelapa Gading	80000	80000	80000	75000	70000	70000	70000	70000	70000	65000	60000	65000	65000	65000	65000	65000	65000	50000
1. Pasar Kramat Jati	80000	70000	70000	80000	70000	70000	65000	65000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
1. Pasar Lenteng Agung	60000	60000	55000	65000	60000	70000	65000	65000	60000	60000	60000	60000	60000	60000	60000	60000	55000	55000
1. Pasar Mampang Prapatan	55000	55000	70000	70000	70000	70000	70000	70000	70000	70000	70000	60000	60000	55000	55000	55000	55000	50000

Cabai Merah Keriting Februari 2020.xls [Compatibility Mode] - Excel

Ibnu Fajar Setiawan

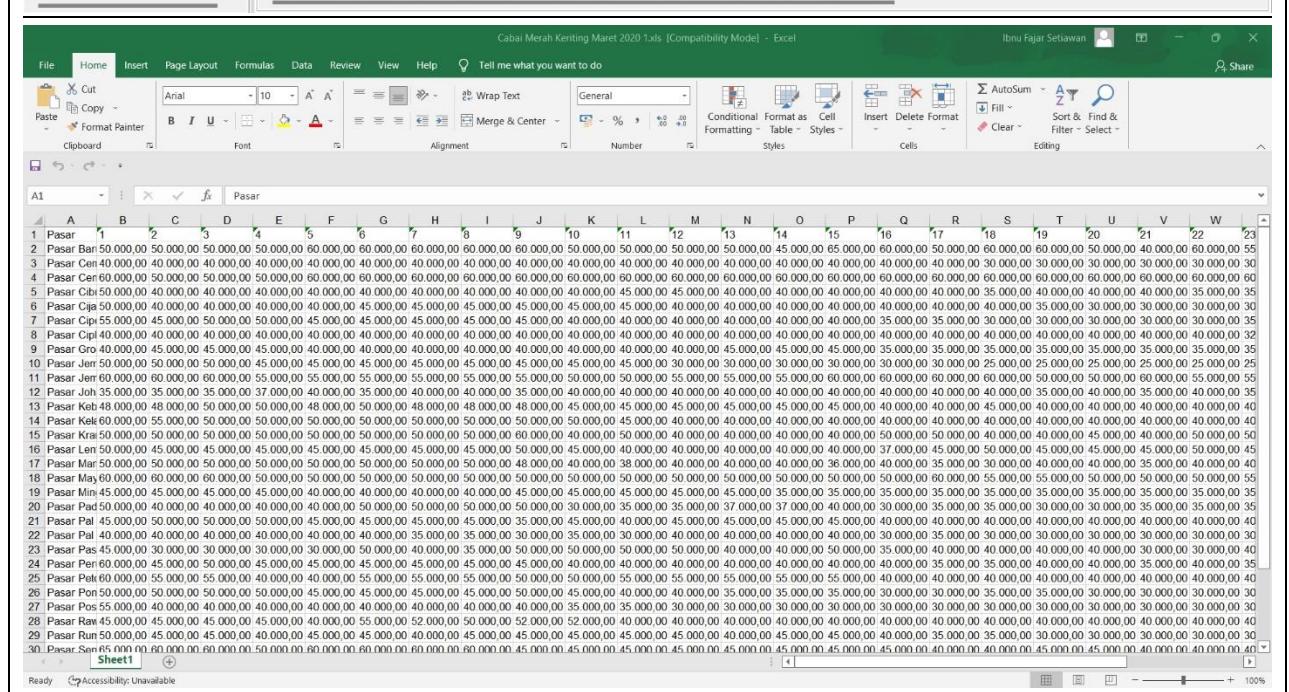
## 3. Cleansing Maret 2020





**Execution Results**

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar																		
1.	Pasar Baru Metro Atom	50000	50000	50000	50000	60000	60000	60000	60000	50000	50000	50000	45000	65000	60000	50000	60000		
2.	Pasar Cempaka Putih	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	30000
3.	Pasar Cengkareng	60000	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
4.	Pasar Clubbur	50000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	35000
5.	Pasar Cijantung	50000	40000	40000	40000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	40000
6.	Pasar Cipete	55000	45000	50000	50000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	35000
7.	Pasar Ciplak	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
8.	Pasar Grogol	40000	45000	45000	45000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	35000
9.	Pasar Jembatan Lima	50000	50000	50000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	35000
10.	Pasar Jembatan Merah	60000	60000	55000	55000	55000	55000	55000	55000	50000	50000	55000	55000	55000	55000	55000	55000	55000	60000
11.	Pasar Johar Baru	35000	35000	35000	37000	40000	35000	40000	40000	35000	40000	40000	40000	40000	40000	40000	40000	40000	40000
12.	Pasar Kebayoran Lama	48000	48000	50000	48000	50000	48000	48000	48000	48000	48000	48000	48000	48000	48000	48000	48000	48000	45000
13.	Pasar Kelapa Gading	60000	55000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	40000
14.	Pasar Kramat Jati	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	40000
15.	Pasar Lenteng Agung	50000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	37000
16.	Pasar Mampang Prapatan	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	35000

#### 4. Cleansing April 2020



Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 Pasar Baru Metro Atom	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	40000	50000	50000	50000	50000
2 Pasar Cempaka Putih	40000	40000	40000	40000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3 Pasar Cengkareng	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	25000
4 Pasar Cibubur	30000	35000	30000	30000	30000	35000	25000	28000	30000	28000	25000	28000	30000	25000	25000	25000	28000	25000
5 Pasar Cijantung	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	25000
6 Pasar Cipete	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	30000
7 Pasar Ciplak	35000	30000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	35000	20000	20000	20000	20000	20000
8 Pasar Grogol	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
9 Pasar Jembatan Lima	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	20000	20000	20000	17000	18000
1. Pasar Jembatan Merah	50000	55000	55000	55000	50000	50000	55000	55000	55000	55000	55000	55000	55000	40000	40000	40000	40000	40000
1. Pasar Johar Baru	30000	30000	35000	25000	25000	25000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	20000
1. Pasar Kebayoran Lama	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
1. Pasar Kelapa Gading	35000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
1.. Pasar Kramat Jati	40000	40000	40000	40000	40000	35000	35000	40000	35000	35000	35000	35000	35000	25000	25000	25000	25000	25000
1.. Pasar Lenteng Agung	45000	45000	45000	45000	40000	35000	37000	30000	30000	30000	30000	30000	30000	30000	30000	30000	28000	28000
1.. Pasar Mampang Prapatan	40000	40000	40000	35000	35000	35000	35000	22000	22000	35000	35000	25000	25000	25000	25000	25000	25000	30000

## 5. Cleansing Mei 2020



The screenshot shows the Apache NiFi interface for a data cleaning process named 'Cleansing Cabai Merah Keriting Mei 2020'. The Transformation Editor on the left lists various components like CSV file input, Select values, Sort rows, Filter rows, and Microsoft Excel output. The Execution Results panel displays a table of data with columns labeled 1 through 18. The data consists of 30 rows, each representing a market ('Pasar') with its name and a series of 18 numerical values. Below this is a screenshot of Microsoft Excel showing the same data in a grid format.

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	30000	35000	35000	35000	30000	40000	40000	40000	40000	35000	40000	35000	35000	35000	35000	35000	35000	35000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cilubur	24000	22000	22000	26000	20000	20000	22000	22000	25000	25000	25000	24000	22000	24000	24000	24000	24000	25000
5	Pasar Cijantung	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
6	Pasar Cipete	25000	25000	25000	20000	20000	25000	25000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
7	Pasar Ciplak	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	23000	22000	22000	22000	22000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
9	Pasar Jembatan Lima	20000	19000	20000	22000	20000	20000	21000	18000	18000	20000	18000	20000	20000	19000	15000	17000	20000	20000
10	Pasar Jembatan Merah	40000	40000	40000	40000	40000	40000	30000	30000	40000	40000	40000	40000	30000	30000	30000	30000	30000	30000
11	Pasar Johar Baru	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
12	Pasar Kebayoran Lama	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
13	Pasar Kelapa Gading	30000	30000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
14	Pasar Lenteng Agung	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	24000	24000	24000	24000	24000
15	Pasar Mampang Prapatan	24000	24000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000

## 6. Cleansing Juni 2020



**Spoon - Cleansing Cabai Merah Keriting Juni 2020**

**Transformations**

- Cleansing Cabai Merah Keriting Juni 2020
  - Run configurations
  - Database connections
  - Steps
  - Hops
  - Partition schemas
  - Slave server
  - Kettle cluster schemas
  - Data Services
  - Hadoop clusters
  - VFS Connections

**Execution Results**

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	50000	40000	40000	40000	35000	40000	40000	35000	35000	35000	35000	35000	35000	35000	40000	40000	35000	
2	Pasar Cempaka Putih	30000	30000	30000	20000	30000	30000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cibubur	20000	20000	20000	25000	20000	20000	22000	20000	20000	20000	20000	22000	22000	22000	20000	20000	22000	20000
5	Pasar Cijantung	25000	25000	25000	25000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
6	Pasar Cipete	25000	25000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
7	Pasar Cipatak	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	18000	18000	18000
8	Pasar Grogol	24000	24000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
9	Pasar Jembatan Lima	18000	17000	15000	15000	14000	15000	15000	15000	15000	15000	15000	14000	14000	15000	15000	17000	15000	15000
10	1. Pasar Jembatan Merah	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	30000	30000	30000	30000
11	1. Pasar Johar Baru	20000	20000	25000	25000	20000	25000	20000	20000	20000	20000	20000	20000	20000	20000	24000	25000	25000	25000
12	1. Pasar Kebayoran Lama	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
13	1. Pasar Kelapa Gading	25000	25000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
14	1. Pasar Kramat Jati	30000	25000	25000	25000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
15	1. Pasar Lenteng Agung	30000	30000	30000	30000	35000	30000	28000	25000	24000	25000	25000	25000	25000	25000	25000	24000	25000	25000
16	1. Pasar Mampang Prapatan	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000

**Cabai Merah Keriting Juni 2020.xls [Compatibility Mode] - Excel**

**Ibnu Fajar Setiawan**

**Execution Results**

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	Pasar Baru Metro Atom	50000	40000	40000	40000	35000	40000	40000	35000	35000	35000	35000	35000	35000	35000	40000	40000	35000	30000
3	Pasar Cempaka Putih	30000	30000	30000	20000	30000	30000	30000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
4	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
5	Pasar Cibarito	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
6	Pasar Ciputat	25000	25000	25000	25000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
7	Pasar Ciputih	25000	25000	25000	25000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
8	Pasar Ciputih	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
9	Pasar Grogol	24000	24000	24000	24000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
10	Pasar Jembatan Lima	18000	17000	15000	15000	14000	15000	15000	15000	15000	15000	15000	14000	14000	15000	15000	16000	16000	16000
11	Pasar Jembatan Merah	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	30000	30000	30000	30000
12	Pasar Johar Baru	20000	20000	25000	25000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
13	Pasar Kebayoran Lama	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
14	Pasar Kramat Jati	30000	25000	25000	25000	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
15	Pasar Lenteng Agung	30000	30000	30000	30000	35000	30000	28000	25000	24000	25000	25000	25000	25000	25000	25000	25000	25000	25000
16	Pasar Mampang Prapatan	40000	40000	40000	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
17	Pasar Pademangan	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
18	Pasar Pakualan	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
19	Pasar Palmerah	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
20	Pasar Pulo Gebang	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
21	Pasar Rawamangun	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
22	Pasar Rawa Pening	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
23	Pasar Senen	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
24	Pasar Tegalrejo	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
25	Pasar Tebet	35000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
26	Pasar Timbulharjo	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
27	Pasar Tomang	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
28	Pasar Tanah Abang	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
29	Pasar Tanah Tinggi	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
30	Pasar Senayan	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000

The screenshot shows the Spoon Data Transformation interface. On the left, the 'Transformations' tree view is expanded to show 'Cleansing Cabai Merah Keriting Juli 2020'. The main workspace displays a flow diagram with five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the workspace is the 'Execution Results' panel, which contains a table of data. The table has columns labeled 1 through 18. The data consists of 30 rows, each starting with 'Pasar' followed by a name and various numerical values.

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar	35000	35000	35000	35000	35000	30000	30000	35000	30000	30000	30000	30000	30000	35000	30000	30000	35000	
2	Pasar Baru Metro Atom	15000	15000	15000	15000	15000	15000	15000	15000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
3	Pasar Cempaka Putih	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
4	Pasar Cengkareng	20000	20000	20000	20000	28000	28000	22000	22000	24000	24000	20000	22000	24000	25000	25000	25000	25000	25000
5	Pasar Cijantung	20000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
6	Pasar Cipete	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	22000	22000	22000	22000	22000
7	Pasar Ciplak	18000	18000	18000	18000	18000	18000	18000	18000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	22000	22000	22000	22000
9	Pasar Jembatan Lima	17000	17000	18000	19000	20000	15000	16000	16000	16000	16000	18000	20000	20000	20000	18000	19000		
10	1. Pasar Jembatan Merah	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	30000	30000	30000	30000
11	1. Pasar Johar Baru	25000	25000	25000	25000	30000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
12	1. Pasar Kebayoran Lama	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
13	1. Pasar Kelapa Gading	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
14	1. Pasar Lenteng Agung	24000	24000	24000	24000	24000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
15	1. Pasar Mampang Prapatan	20000	20000	20000	30000	30000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	24000

The screenshot shows a Microsoft Excel spreadsheet titled 'Cabai Merah Keriting Juli 2020.xls'. The visible portion of the sheet contains data from columns A to W. The data appears to be a subset of the cleansing results, showing names like 'Pasar' and 'Pasar Baru Metro Atom' followed by numerous zeros. The Excel ribbon is visible at the top, and the formula bar shows 'Pasar'.

## 8. Cleansing Agustus 2020



The screenshot shows the Apache Kettle Transformation Designer interface. The left pane displays the 'Transformations' tree, with 'Cleansing Cabai Merah Kriting Agustus' selected. The right pane shows the 'Execution Results' table, which contains 30 rows of data. The table has columns labeled 1 through 18, representing various fields. The data includes entries such as 'Pasar Baru Metro Atom', 'Pasar Cempaka Putih', and 'Pasar Kelapa Gading'. The table also includes several entries starting with '1.' followed by names like 'Jembatan Lima', 'Jembatan Merah', 'Johar Baru', etc.

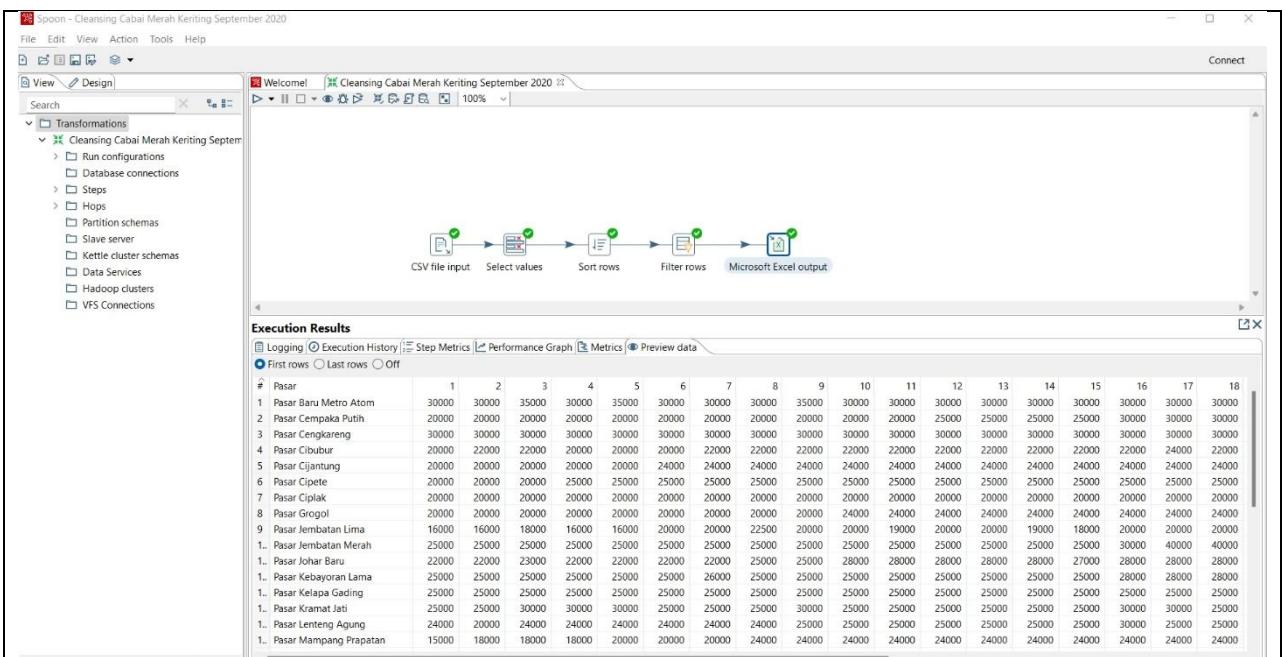
#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	30000	30000	30000	35000	35000	30000	30000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	30000
2	Pasar Cempaka Putih	22000	22000	22000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000
3	Pasar Cengkareng	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	30000
4	Pasar Cibubur	28000	28000	28000	22000	22000	28000	24000	25000	25000	26000	24000	22000	24000	24000	24000	24000	24000	22000
5	Pasar Cijantung	28000	25000	28000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000
6	Pasar Cipete	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	20000
7	Pasar Ciplik	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	22000	22000	22000	22000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	22000
9	Pasar Jembatan Lima	20000	21000	25000	20000	18000	18000	18000	18000	18000	19000	19000	18000	18000	18000	18000	18000	16000	17000
10	Pasar Jembatan Merah	30000	30000	40000	40000	40000	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	30000
11	Pasar Johar Baru	27000	27000	26000	27000	23000	24000	24000	24000	24000	24000	26000	26000	35000	23000	24000	25000	26000	26000
12	Pasar Kebayoran Lama	25000	25000	25000	28000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
13	Pasar Kelapa Gading	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000	25000
14	Pasar Kramat Jati	25000	25000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	25000
15	Pasar Lenteng Angin	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
16	Pasar Mampang Prapatan	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000	24000

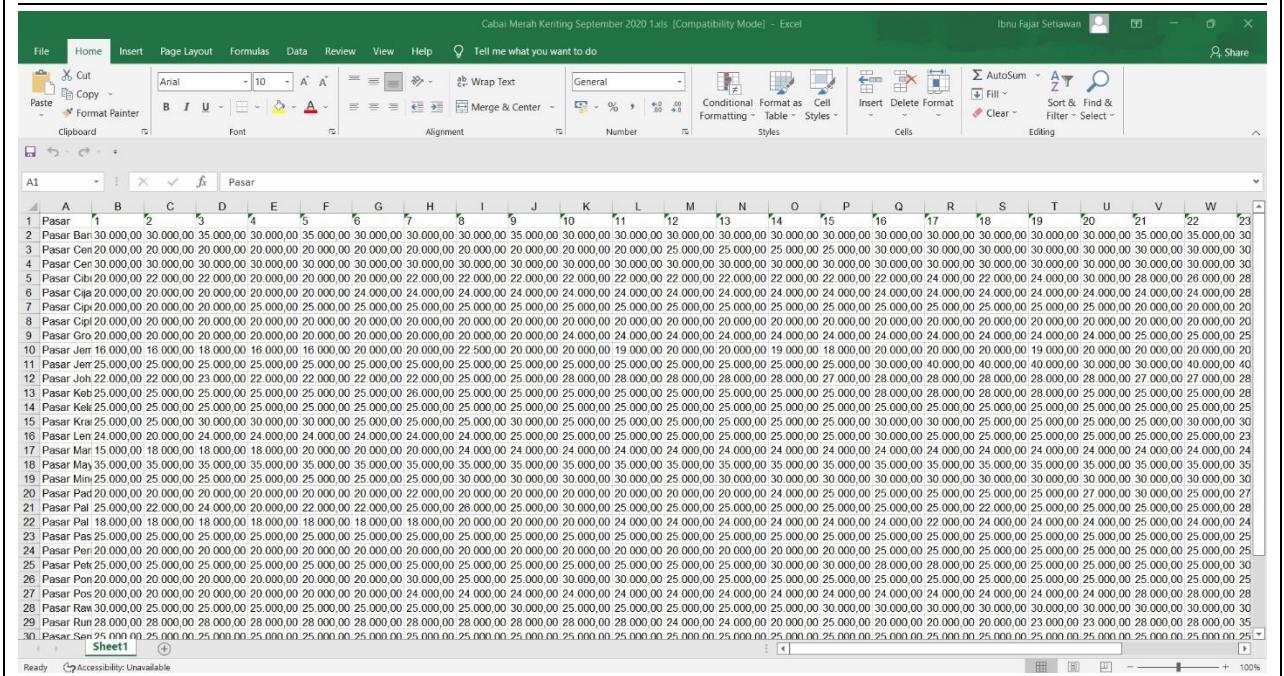
The screenshot shows a Microsoft Excel spreadsheet titled 'Cabai Merah Kriting Agustus 2020.xls'. The table structure is identical to the one in the Spoon interface, containing 30 rows of data from A1 to W23. The data includes various market names and their corresponding values across 18 columns. The Excel interface shows the ribbon, toolbars, and standard cell editing features.

## 9. Cleansing September 2020





The screenshot shows the Spoon interface for Apache Nifi. On the left, the 'Transformations' tree view is expanded, showing a configuration named 'Cleansing Cabai Merah Keriting Septem'. The main workspace displays a flow with five components: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the workspace is the 'Execution Results' panel, which lists the first 30 rows of data. The data consists of 18 columns, with the first column labeled '# Pasar' and subsequent columns labeled 1 through 18. The data includes various market names like Pasar Baru Metro Atom, Pasar Cempaka Putih, etc., along with their respective values.

The screenshot shows a Microsoft Excel spreadsheet titled 'Cabai Merah Keriting September 2020.xls'. The data is identical to the one shown in the Spoon interface, spanning from row 1 to row 30. The columns are labeled A through W, and the data includes market names and their corresponding values across 18 columns.

## 10. Cleansing Oktober 2020



The screenshot shows the Apache Kettle Transformation Editor. The left pane displays the 'Transformations' tree with a node named 'Cleansing Cabai Merah Keriting Oktober'. The right pane shows a flow diagram with five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the flow is the 'Execution Results' table, which contains data for 30 different markets (Pasar) across 18 columns. The table includes columns for market names like Pasar Baru Metro Atom, Pasar Cempaka Putih, etc., and numerical values ranging from 30,000 to 40,000.

**Cleansing Cabai Merah Keriting Oktober**

Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Pasar	40000	35000	35000	35000	45000	45000	45000	45000	45000	40000	45000	45000	50000	50000	50000	50000	50000	50000
2. Pasar Baru Metro Atom	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3. Pasar Cempaka Putih	30000	30000	30000	30000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
4. Pasar Cengkareng	28000	28000	35000	30000	40000	38000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
5. Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
6. Pasar Cipete	25000	25000	25000	28000	38000	40000	40000	40000	40000	42000	36000	40000	40000	40000	40000	40000	40000	40000
7. Pasar Cipatak	20000	20000	20000	20000	20000	20000	20000	20000	20000	28000	30000	38000	38000	38000	38000	38000	38000	38000
8. Pasar Grogol	30000	30000	30000	38000	38000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
9. Pasar Jembatan Lima	25000	25000	30000	30000	32000	30000	30000	35000	30000	30000	32000	30000	30000	30000	30000	30000	30000	30000
10. Pasar Jembatan Merah	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
11. Pasar Johar Baru	30000	30000	30000	40000	40000	35000	35000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000
12. Pasar Kebayoran Lama	40000	38000	40000	38000	39000	39000	39000	38000	38000	39000	39000	40000	40000	40000	40000	40000	40000	40000
13. Pasar Kelapa Gading	25000	25000	25000	35000	40000	40000	40000	40000	40000	40000	40000	45000	45000	45000	45000	45000	45000	45000
14. Pasar Kramat Jati	45000	40000	40000	30000	30000	45000	45000	40000	45000	45000	45000	40000	45000	45000	40000	40000	40000	40000
15. Pasar Lenteng Agung	30000	30000	30000	30000	35000	35000	45000	45000	45000	45000	45000	42000	40000	40000	40000	40000	40000	40000
16. Pasar Mampang Prapatan	34000	38000	38000	34000	34000	38000	40000	40000	40000	35000	35000	38000	38000	38000	38000	38000	38000	38000

**Cabai Merah Keriting Oktober 2020.xls [Compatibility Mode] - Excel**

The screenshot shows an Excel spreadsheet titled 'Cabai Merah Keriting Oktober 2020.xls'. The table structure is identical to the one in the Spoon interface, containing 30 rows of market data and 18 columns of numerical values. The Excel ribbon and various toolbars are visible at the top.

## 11. Cleansing November 2020



The screenshot shows the Apache Nifi interface for data cleansing. The top part displays a flow diagram with five main components: CSV file input, Select values, Sort rows, Filter rows, and Microsoft Excel output. The 'Transformations' tree on the left shows a single configuration named 'Cleansing Cabai Merah Keriting Novem'. The 'Execution Results' table below contains 30 rows of data, each representing a market (Pasar) with various numerical values. The bottom part shows the resulting data in Microsoft Excel format, with the first few rows of data visible.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar																	
2	Pasar Baru Metro Atom	50000	50000	50000	50000	50000	50000	50000	50000	50000	45000	45000	50000	50000	50000	45000	45000	45000
3	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	40000	40000	40000	40000	40000	40000	40000	40000	40000
4	Pasar Cengkareng	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
5	Pasar Cibubur	45000	40000	38000	40000	38000	30000	30000	40000	35000	35000	40000	40000	40000	40000	40000	40000	40000
6	Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000
7	Pasar Cipete	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
8	Pasar Ciplak	36000	40000	38000	36000	36000	36000	36000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
9	Pasar Grogol	40000	40000	40000	40000	35000	35000	35000	38000	38000	38000	38000	38000	38000	38000	38000	38000	38000
10	Pasar Jembatan Lima	33000	33000	32000	32000	32000	31000	30000	29000	31000	30000	28000	28000	28000	28000	30000	31000	32000
11	1. Pasar Jembatan Merah	40000	40000	40000	50000	50000	45000	45000	45000	45000	50000	50000	40000	40000	40000	45000	50000	50000
12	1. Pasar Johar Baru	42000	42000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
13	1. Pasar Kebayoran Lama	40000	45000	45000	45000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	45000	45000	40000
14	1. Pasar Kelapa Gading	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000
15	1. Pasar Kramat Jati	50000	55000	55000	40000	45000	45000	45000	45000	40000	40000	40000	40000	40000	40000	50000	45000	45000
16	1. Pasar Lenteng Agung	45000	40000	45000	45000	40000	45000	45000	45000	45000	40000	40000	40000	40000	40000	45000	45000	40000
17	1. Pasar Mampang Prapatan	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000	44000

## 12. Cleansing Desember 2020



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar																	
2	Pasar Baru Metro Atom	45000	45000	55000	60000	60000	65000	65000	65000	65000	65000	65000	65000	65000	60000	60000	65000	
3	Pasar Cempaka Putih	45000	35000	45000	45000	45000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	60000
4	Pasar Cengkareng	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	60000
5	Pasar Cijantung	45000	45000	45000	45000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	65000
6	Pasar Cipete	45000	45000	46000	46000	46000	55000	55000	55000	55000	55000	55000	55000	55000	58000	50000	50000	55000
7	Pasar Cipatik	35000	35000	35000	35000	40000	40000	35000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
8	Pasar Grogol	38000	38000	48000	48000	48000	48000	54000	53000	53000	53000	53000	53000	50000	50000	50000	60000	63000
9	Pasar Jembatan Lima	44000	45000	43000	43000	47000	45000	46000	49000	50000	45000	47000	45000	42000	45000	43000	45000	55000
10	1. Pasar Jembatan Merah	50000	50000	50000	60000	60000	70000	70000	65000	65000	65000	65000	65000	65000	65000	60000	60000	60000
11	1. Pasar Johor Baru	50000	50000	45000	50000	50000	50000	50000	50000	50000	50000	50000	50000	50000	43000	45000	50000	60000
12	1. Pasar Kebayoran Lama	50000	50000	50000	45000	45000	40000	60000	50000	48000	60000	60000	60000	58000	58000	60000	55000	65000
13	1. Pasar Kelapa Gading	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000	55000
14	1. Pasar Kramat Jati	45000	45000	50000	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	55000	55000	55000	70000
15	1. Pasar Lenteng Agung	45000	45000	45000	45000	45000	45000	45000	45000	45000	60000	60000	60000	60000	60000	60000	60000	60000
16	1. Pasar Mampang Prapatan	44000	50000	50000	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	65000

#### 4. File Praktikum

Github Repository:

<https://github.com/IbnuFajar7/Data-Warehouse/tree/main/Prak-11>



## 5. Soal Latihan

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:

1. Data cleansing adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat.
2. Menghilangkan kesalahan dan inkonsistensi yang muncul saat beberapa data sources dikumpulkan dalam satu dataset. Meningkatkan efisiensi kerja karena proses ini akan memudahkan Anda dan tim pengolah data untuk menemukan apa yang dibutuhkan dari data.

## 6. Kesimpulan

- a. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui bahwa kita bisa melakukan cleansing data yang bisa memudahkan kita untuk cek data yang kemungkinan kita salah input, dengan metode ini kita akan bisa lebih mudah tanpa harus cek satu-satu dan memakan waktu yang lebih lama, waktu ini akan membuat kita lebih efisien

## 7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Tugas Akhir Praktikum III	✓	

## 8. Formulir Umpam Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Tugas Akhir Praktikum III	30 Menit	1

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

